University of Cyprus
Department of Economics

Working Paper 06-2022

Strictly log-concave probability distributions in discrete response models

Eleni Aristodemou

Department of Economics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
Web site: http://www.ucy.ac.cy/econ/en

Strictly log-concave probability distributions in discrete response models*

Eleni Aristodemou†

University of Cyprus

July 5, 2022

Abstract

This paper extends the results of Prékopa (1973, 1980) on strictly log-concave cumulative distributions to strictly log-concave probability distributions. It is shown that if a random variable follows a strictly logarithmic concave distribution, then the probability that the random variable is contained within a convex polytope is also strictly logarithmic concave. This formal result can be useful for identification and estimation of a general class of linear-index discrete response models, where the additively separable unobservable follows a strictly log-concave distribution.

Keywords: Discrete Response Models, Strict Log-Concavity

*I would like to thank the anonymous referee for the insightful suggestions and comments. I would also like to thank Andrew Chesher, Lars Nesheim, Adam Rosen and Frank Windmeijer for helpful discussions and comments. Eletheria Kelekidou provided excellent research assistance. I gratefully acknowledge financial support from ESRC and UCL. Any errors are my own.

†Address: Department of Economics, University of Cyprus, P.O.Box 20537, CY-1678 Nicosia, Cyprus. Email: aristodemou.d.eleni@ucy.ac.cy

1 Introduction

In discrete response threshold crossing models, where the dependent variable takes values from a finite set of discrete outcomes, the strict log-concavity of the likelihood function in the parameters of interest plays a crucial role in their identification and estimation. The literature has shown that if the likelihood function is strictly log-concave in the parameters of interest, then unique estimation of those parameters in the binary and multinomial discrete response models can be achieved[1].

In the most commonly used form of discrete response models, where the disturbance term is additively separable, the log-concavity of the likelihood function is directly related to the log-concavity of the probability density function of the random unobservables and many widely used distributions are known to be log-concave (Bagnoli and Bergstrom (2005)). This paper extends the work of Prékopa (1973, 1980) who showed that the cumulative distribution function (CDF) of a random variable with a strictly log-concave probability density function (PDF) is also strictly log-concave, to show that the probability *distribution* function is also strictly log-concave. In particular, it is shown that if a random variable has a strictly logarithmic concave density, then the probability that it is contained in a convex polytope is also strictly log-concave. This result is of particular importance when examining linear (single and multiple) index discrete response models.

The contribution of this paper is two-fold. It lays out a general formalization of some well known results in the literature, related for example to the classic probit and logit models that, to the best of our knowledge, have not been explicitly shown, and formally states a result applicable to a wider class of linear index discrete response models with additively separable disturbance term, whose distribution is strictly log-concave. For example in economic applications that model the demand for differentiated products, the choice individuals make can be multidimensional and the choice probabilities are given by the probability the random vector

[1] See for example Albert and Anderson (1984), Delle Site et al. (2019), Demidenko (2001), Lesaffre and Kaufmann (1992), Mäkeläinen et al. (1981), McCullagh (1980), Orme and Ruud (2002), Pratt (1981), Silvapulle (1981) and Silvapulle and Burridge (1986).

of unobservables is contained in a convex polytope. If these multidimensional unobservables jointly follow a strictly log-concave distribution, then the probability they are jointly contained in such a polytope is strictly log-concave in the thresholds that define the half-space representation of the convex polytope. Since these thresholds are linear in the parameters, under the standard rank condition as discussed in Wedderburn (1976), the probabilities are guaranteed to be strictly log-concave in the model parameters, too. This provides a sufficient condition for point-identification and unique estimation of the regression parameters.

1.1 Notation Section

For the rest of the paper the following notations are used. $V \in \mathbb{R}^n$ denotes a random vector, with PDF, f. Throughout, f is assumed to be log-concave in V, that is $\forall \lambda \in (0,1)$ and $\forall\, V_1, V_2 \in \mathbb{R}^n$,

$$f(\lambda V_1 + (1-\lambda)V_2) \geq f(V_1)^{\lambda} f(V_2)^{1-\lambda}.$$

When the above inequality is strict f is strictly log-concave in V.

The probability that the random vector, $V \in \mathbb{R}^n$, is contained in a convex polytope A is denoted by $P(A) = P(V \in A)$. A convex polytope is a convex set of points in $\mathbb{R}^n$ defined by the intersection of a finite number of half-spaces. The half-space or H-representation of a convex polytope A is given by,

$$A = \{HV \leq \ell\}$$

where H is a $(q \times n)$ matrix, V is a $(n \times 1)$ vector and ℓ is a $(q \times 1)$ vector of constant thresholds.

Define two convex polytopes A and B, then the Minkowski addition of those sets is given by,

$$A + B = \{a + b | a \in A, b \in B\} \tag{1}$$

2 Strictly Log-concave Probability Distribution Functions

Similarly to Section 1.1, a function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}_+$ is said to be strictly logarithmic concave on $\mathbb{R}^n$ if for any pair of $V_1, V_2 \in \mathbb{R}^n$ and for every $0 < \lambda < 1$,

$$\mathcal{F}(\lambda V_1 + (1-\lambda)V_2) > \mathcal{F}(V_1)^{\lambda}\mathcal{F}(V_2)^{1-\lambda}.$$

Theorem 4 of Prékopa (1973) provides a set of four conditions which if satisfied the measure $\mathcal{P}$ defined on measurable subsets of $\mathbb{R}^n$ and generated by the strictly logarithmic concave function $\mathcal{F}$ is also strictly logarithmic concave , i.e.

$$\mathcal{P}(\lambda A + (1-\lambda)B) > \mathcal{P}(A)^{\lambda}\mathcal{P}(B)^{1-\lambda}$$

where A and B are convex subsets of $\mathbb{R}^n$ with the property that $0 < \mathcal{P}(A) < \infty$ and $0 < \mathcal{P}(B) < \infty$.

Theorem 5 of Prékopa (1973) uses the result in Theorem 4 to show that, if the PDF of a random vector $V \in \mathbb{R}^n$, f, is positive and strictly log-concave in an open convex set D then its corresponding CDF, F [2], is also strictly logarithmic concave in the set D.

This section extends Theorem 5 of Prékopa (1973) to show that if f is positive and strictly log-concave in an open convex set D, then the measure P, the probability *distribution* function generated by f, is also strictly logarithmic concave in the set D, where P is given by,

$$P(S) = P(V \in S) \tag{2}$$

and S is a measurable set in $\mathbb{R}^n$.

Such an extension is useful in proving point-identification and unique estimation in many discrete response models where the additively separable unobservables follow a strictly log-concave distribution and the probability of observing a specific outcome is given by the form in (2).

[2] Theorem 5 of Prékopa (1973) calls F, defined in its proof as $F(u) = P(A)$, where $A = \{x | x \leq u\}$, $x \in \mathbb{R}^n$, the probability distribution function, instead of the cumulative distribution function. See Prékopa (2007) Theorem 4, Prékopa (2012) Theorem 3, and Delle Site et al. (2019) page 104.

Theorem 1. *If f is a positive and strictly log-concave density function in an open convex set $D \subset \mathbb{R}^n$ then the probability distribution function generated by f, P, is strictly log-concave in this open convex set D.*

Proof. The proof follows similar arguments as the proof of Theorem 5 in Prékopa (1973) on pages 7-8. Define two convex sets A and B in the interior of $D \subset \mathbb{R}^n$ with $A \neq B$ as

$$\begin{aligned} A &= \{V | HV \leq \ell_1\} \\ B &= \{V | HV \leq \ell_2\} \end{aligned} \tag{3}$$

where H is a $(q \times n)$ matrix, V is a $(n \times 1)$ vector and ℓ_1 and ℓ_2 are $(q \times 1)$ vectors of constants with $\ell_1 \neq \ell_2$[3], such that $0 < P(A) < \infty$ and $0 < P(B) < \infty$. Define the vectors of the k upper bounds of A and B as $\ell_1^u = (\ell_{11}, \ldots, \ell_{1k})$ and $\ell_2^u = (\ell_{21}, \ldots, \ell_{2k})$, respectively, and fix $\ell_1^u < \ell_2^u$. Following Theorem 4 of Prékopa (1973), thereafter Prékopa4, if for every $\lambda \in (0, 1)$ the two sets can be decomposed as $A_1 \cup A_2 = A$, $A_1 \cap A_2 = \emptyset$, $B_1 \cup B_2 = B$, $B_1 \cap B_2 = \emptyset$ and $A_1 \cap B_1 = \emptyset$[4], so that the four conditions given below are satisfied, then it can be shown that

$$P[V \in (\lambda A + (1 - \lambda)B)] > [P(V \in A)]^{\lambda}[P(V \in B)]^{1-\lambda}.$$

Define A_1, B_1, A_2 and B_2 as

$$A_1 = \left\{ V \middle| \; HV \leq \ell_1, \sum_{i=1}^{k}(K_i V) \geq \sum_{i=1}^{k} \ell_{1i} - \delta \right\} \tag{4}$$

$$B_1 = \left\{ V \middle| \; HV \leq \ell_2, \sum_{i=1}^{k}(K_i V) \geq \sum_{i=1}^{k} \ell_{2i} - \zeta \right\} \tag{5}$$

$$A_2 = A - A_1, B_2 = B - B_1 \tag{6}$$

such that

$$A_2 : \left\{ V \middle| \; HV \leq \ell_1, \sum_{i=1}^{k}(K_i V) < \sum_{i=1}^{k} \ell_{1i} - \delta \right\} \tag{7}$$

[3]The two convex sets are in fact convex polytopes made from the intersection of a finite number of half-spaces, linear in $V \in \mathbb{R}^n$. This linearity allows for linear combinations of half-spaces to be chosen which makes the proof more tractable.

[4]The proof of Theorem 4 of Prékopa (1973) on page 6 requires the sets A_1 and B_1 to be disjoint sets.

$$B_2 : \left\{ V \middle| \; HV \leq \ell_2, \sum_{i=1}^{k}(K_i V) < \sum_{i=1}^{k} \ell_{2i} - \zeta \; \right\} \tag{8}$$

where $\delta > 0$ and $\zeta > 0$, such that $A_2 \neq \emptyset$, $B_2 \neq \emptyset$ and $A_1 \cap B_1 = \emptyset$. K is a $k \times n$ matrix with $k \leq q$ such that, $K_i V = \ell_{1i}$ and $K_i V = \ell_{2i}$ correspond to the i^{th} upper hyperplane of A and B, respectively.

Condition 1 (Prékopa4): *A_1 and B_1 are bounded, closed and convex sets, and A_2 and B_2 are convex sets.*

Suppose that $\hat{V} \in A_1$ and $\tilde{V} \in A_1$ and define $V^1 = \lambda \hat{V} + (1-\lambda)\tilde{V}$, where $\lambda \in (0,1)$. Using (4), notice that:

$$\begin{aligned} \lambda H \hat{V} &\leq \lambda \ell_1 \\ (1-\lambda) H \tilde{V} &\leq (1-\lambda)\ell_1 \end{aligned}$$

$\Rightarrow$

$$\lambda H \hat{V} + (1-\lambda) H \tilde{V} \leq \lambda \ell_1 + (1-\lambda)\ell_1$$

$\Rightarrow$

$$H[\lambda \hat{V} + (1-\lambda)\tilde{V}] \leq \ell_1$$

$\Rightarrow$

$$HV^1 \leq \ell_1 \tag{9}$$

and that,

$$\begin{aligned} \lambda \sum_{i=1}^{k}(K_i \hat{V}) &\geq \lambda \sum_{i=1}^{k} \ell_{1i} - \lambda\delta \\ (1-\lambda) \sum_{i=1}^{k}(K_i \tilde{V}) &\geq (1-\lambda) \sum_{i=1}^{k} \ell_{1i} - (1-\lambda)\delta \end{aligned}$$

$\Rightarrow$

$$\lambda \sum_{i=1}^{k}(K_i \hat{V}) + (1-\lambda) \sum_{i=1}^{k}(K_i \tilde{V}) \geq \sum_{i=1}^{k} \ell_{1i} - \delta$$

$\Rightarrow$

$$\sum_{i=1}^{k}[\lambda(K_i \hat{V}) + (1-\lambda)(K_i \tilde{V})] \geq \sum_{i=1}^{k} \ell_{1i} - \delta$$

$\Rightarrow$

$$\sum_{i=1}^{k}[K_i(\lambda \hat{V} + (1-\lambda)\tilde{V})] \geq \sum_{i=1}^{k} \ell_{1i} - \delta$$

$\Rightarrow$

$$\sum_{i=1}^{k}(K_iV^1) \geq \sum_{i=1}^{k}\ell_{1i} - \delta \tag{10}$$

Combining (9) and (10),

$$V^1 : HV^1 \leq \ell_1, \sum_{i=1}^{k}(K_iV^1) \geq \sum_{i=1}^{k}\ell_{1i} - \delta$$

$\therefore V^1 \in A_1$, therefore A_1 is convex. The fact that A_1 is closed and bounded, follows from the weak inequalities in (4). Following similar arguments, it can be shown that B_1 is closed, bounded, and convex, while A_2 and B_2 are convex sets.

Condition 2 (Prékopa4): *The following relations hold:*

$$\lambda A_1 + (1-\lambda)B_1 \quad \cup \quad \lambda A_2 + (1-\lambda)B_2 = \lambda A + (1-\lambda)B \tag{11}$$

$$\lambda A_1 + (1-\lambda)B_1 \quad \cap \quad \lambda A_2 + (1-\lambda)B_2 = \emptyset \tag{12}$$

$$\lambda A_1 + (1-\lambda)B_1 \quad \cap \quad A_1 = \emptyset \tag{13}$$

$$\lambda A_1 + (1-\lambda)B_1 \quad \cap \quad B_1 = \emptyset \tag{14}$$

Definitions (4), (7), (5), and (8) imply that for any fixed $\lambda \in (0,1)$, $\lambda A_1 + (1-\lambda)B_1$ can be expressed as,

$$\left\{V \middle| \; HV \leq \lambda\ell_1 + (1-\lambda)\ell_2, \lambda\sum_{i=1}^{k}(K_iV) + (1-\lambda)\sum_{i=1}^{k}(K_iV) \geq \sum_{i=1}^{k}(\lambda\ell_{1i} + (1-\lambda)\ell_{2i}) - \lambda\delta - (1-\lambda)\zeta \right\}$$

$\Rightarrow$

$$\left\{V \middle| \; HV \leq \lambda\ell_1 + (1-\lambda)\ell_2, \sum_{i=1}^{k}(K_iV) \geq \sum_{i=1}^{k}(\lambda\ell_{1i} + (1-\lambda)\ell_{2i}) - \lambda\delta - (1-\lambda)\zeta \right\} \tag{15}$$

and $\lambda A_2 + (1-\lambda)B_2$ as

$$\left\{V \middle| \; HV \leq \lambda\ell_1 + (1-\lambda)\ell_2, \sum_{i=1}^{k}(K_iV) < \sum_{i=1}^{k}(\lambda\ell_{1i} + (1-\lambda)\ell_{2i}) - \lambda\delta - (1-\lambda)\zeta \right\} \tag{16}$$

Then following (15) and (16), relations (11) and (12) hold for all δ and ζ.

To prove (13) and (14), first notice that the definitions of A_1 and B_1 imply that the set A_1 is bounded below by the hyperplane $\sum_{i=1}^{k}(K_iV) = \sum_{i=1}^{k}\ell_{1i} - \delta$ and B_1 is bounded below by the hyperplane $\sum_{i=1}^{k}(K_iV) = \sum_{i=1}^{k}\ell_{2i} - \zeta$. Then, proving relations (13) and (14) is equivalent to showing that for any fixed $\lambda \in (0,1)$, there exists a $\zeta > 0$ and a $\delta > 0$ such that the lower bound of $\lambda A_1 + (1-\lambda)B_1$ is above the upper bound of A_1 and the upper bound of $\lambda A_1 + (1-\lambda)B_1$ is below the lower bound of B_1. The latter relation is satisfied if,

$$(\delta, \zeta) : \left\{ \begin{array}{c} \sum_{i=1}^{k} \ell_{2i} - \zeta > \sum_{i=1}^{k} (\lambda \ell_{1i} + (1-\lambda)\ell_{2i}) \\ \lambda \left(\sum_{i=1}^{k} \ell_{2i} \right) - \zeta > \lambda \sum_{i=1}^{k} \ell_{1i} \\ \lambda \left(\sum_{i=1}^{k} \ell_{2i} - \sum_{i=1}^{k} \ell_{1i} \right) > \zeta \end{array} \right\} \tag{17}$$

Since $\ell_1^u < \ell_2^u$, $\left(\sum_{i=1}^{k} \ell_{2i} - \sum_{i=1}^{k} \ell_{1i} \right) > 0$, which implies that there exists $\zeta > 0$ sufficiently small such that (14) holds.

Relation (13) holds for all δ and ζ such that,

$$(\delta, \zeta) : \left\{ \begin{array}{c} \sum_{i=1}^{k} (\lambda \ell_{1i} + (1-\lambda)\ell_{2i}) - \lambda\delta - (1-\lambda)\zeta > \sum_{i=1}^{k} \ell_{1i} \\ \sum_{i=1}^{k} (1-\lambda)\ell_{2i} - (1-\lambda)\zeta > \sum_{i=1}^{k} (\ell_{1i} - \lambda \ell_{1i}) + \lambda\delta \\ \sum_{i=1}^{k} (1-\lambda)\ell_{2i} - (1-\lambda)\zeta > \sum_{i=1}^{k} (1-\lambda)\ell_{1i} + \lambda\delta \\ (1-\lambda) \left[\sum_{i=1}^{k} (\ell_{2i} - \ell_{1i}) - \zeta \right] > \lambda\delta \\ \delta < \frac{(1-\lambda)}{\lambda} \left[\sum_{i=1}^{k} (\ell_{2i} - \ell_{1i}) - \zeta \right] \end{array} \right\} \tag{18}$$

Since $\ell_1^u < \ell_2^u$ and (17) shows that a sufficiently small $\zeta > 0$ exists such that $\left[\sum_{i=1}^{k} (\ell_{2i} - \ell_{1i}) - \zeta \right] > 0$, then a $\delta > 0$ sufficiently small exists.

Condition 3 (Prékopa4): *For the measures of the decomposing sets the following relations hold*

$$P(A_1) > 0 \; ; \; P(B_1) > 0 \tag{19}$$

and

$$\frac{P(A_2)}{P(A_1)} = \frac{P(B_2)}{P(B_1)} \tag{20}$$

Condition 3 (Prékopa4) can been shown to hold following the proof of Theorem 5 Prékopa (1973) on page 8, the steps of which, with a slight change in notation, are included for completeness. Fix δ_0 and ζ_0 to satisfy conditions (13) and (14). Condition (19) is satisfied for any $\delta > 0$ and $\zeta > 0$. If (20) is also satisfied for δ_0 and ζ_0 then the proof is completed. If (20) is not satisfied for δ_0 and ζ_0 since $P(A_1)$ is continuous in δ and $P(B_1)$ is continuous in ζ and

$$\lim_{\delta \to 0} P(A_1) = \lim_{\zeta \to 0} P(B_1) = 0$$

a δ_1 can be found, with $\delta_1 \leq \delta_0$, and a ζ_1, with $\zeta_1 \leq \zeta_0$, to satisfy both relations and (20) is satisfied with these.

Condition 4 (Prékopa4): *f is strictly logarithmic concave in the convex hull $A_1 \cup B_1 \subset D$.*

f is strictly logarithmic concave in $D \subset \mathbb{R}^n$, therefore Condition 4 (Prékopa4) is satisfied. $\square$

Following Theorem 1, Theorem 9 of Prékopa (1980) can also be extended to strictly logarithmic concave functions. Theorem 9 of Prékopa (1980) states that (with a slight change in notation),

"If $g_1(\ell, W), ..., g_r(\ell, W)$ are concave functions in $\mathbb{R}^{q+n}$, where ℓ is a q-component and W is a n-component vector, and V is a n-component random vector whose probability distribution is logarithmic concave in $\mathbb{R}^n$, then the function

$$h(\ell) = P(g_1(\ell, V) \geq 0, \ldots, g_r(\ell, V) \geq 0)$$

is logarithmic concave on $\mathbb{R}^q$".

Theorem 2 below extends Theorem 9 of Prékopa (1980) to strictly log-concave probability measures over convex polytopes. In particular, it is shown that the measure P is strictly logarithmic concave in the thresholds defining the half-space representation of the convex polytope.

Theorem 2. *If $g_1(\ell, W), ..., g_r(\ell, W)$ are linear functions in $\mathbb{R}^{q+n}$, where ℓ is a q-component of constant thresholds and W is a n-component vector, and V is a n-component random vector whose probability density function is strictly logarithmic concave in $\mathbb{R}^n$ then the function*

$$h(\ell) = P(g_1(\ell, V) \geq 0, \ldots, g_r(\ell, V) \geq 0)$$

is strictly logarithmic concave on $\mathbb{R}^q$.

Proof. Define the sets $H(\ell_1)$ and $H(\ell_2)$ as

$$H(\ell_1) \equiv \left\{ W : \begin{array}{c} g_1(\ell_1, W) \geq 0 \\ . \\ \wedge \\ . \\ g_r(\ell_1, W) \geq 0 \end{array} \right\} \quad \textit{and} \quad H(\ell_2) \equiv \left\{ W : \begin{array}{c} g_1(\ell_2, W) \geq 0 \\ . \\ \wedge \\ . \\ g_r(\ell_2, W) \geq 0 \end{array} \right\}$$

where g_j with $j = \{1, \dots, r\}$ are concave functions. Equivalently,

$$\begin{aligned} H(\ell_1) &= \{W | g_j(\ell_1, W) \geq 0, j = 1, \dots, r\} \\ H(\ell_2) &= \{W | g_j(\ell_2, W) \geq 0, j = 1, \dots, r\} \end{aligned}$$

From the proof of Theorem 9 of Prékopa (1980), on pages 13-14, it follows that the set $H = \{\ell | H(\ell) \neq \emptyset\}$, with $\ell_1 \in H(\ell)$ and $\ell_2 \in H(\ell)$, is convex, and if $H \neq \emptyset$, the family $H(\ell)$ is concave on H. Further restricting the functions g_j to be linear functions in $\mathbb{R}^{q+n}$ implies that the set $H = \{\ell | H(\ell) \neq \emptyset\}$ is a convex polytope. Define by $h(\ell) = P[V \in H(\ell)]$ then,

$$\begin{aligned} h(\lambda\ell_1 + (1-\lambda)\ell_2) &= P[V \in H(\lambda\ell_1 + (1-\lambda)\ell_2)] \\ &\geq P[V \in (\lambda H(\ell_1) + (1-\lambda)H(\ell_2))] \\ &> [P(V \in H(\ell_1))]^{\lambda}[P(V \in H(\ell_2))]^{1-\lambda} \\ &= h(\ell_1)^{\lambda} h(\ell_2)^{1-\lambda} \end{aligned} \tag{21}$$

where the strict inequality follows from Theorem 1. This completes the proof that $h(\ell)$ is strictly logarithmic concave in ℓ. □

2.1 Examples

For illustration purposes, Theorems 1 and 2 are applied to two discrete response models with additively separable disturbance terms, the binary response probit model and the bivariate ordered response probit model.

2.1.1 Binary Response Probit Model

Consider the simple binary response probit model,

$$Y = 1(X\beta + V > 0) \tag{22}$$

where $1(\cdot)$ is the indicator function which equals to 1 if $(\cdot)$ is true and 0 otherwise, X is an $(1 \times k)$ vector of regressors, β is a $(k \times 1)$ vector of parameters of interest and $V \in \mathbb{R}$ is the scalar unobserved heterogeneity, such that $V \perp X$ and $V \sim N(0,1)$[5]. Using the notation in Section 2 define as,

$$H^0(\ell) = \{W | g^0(\ell, W) \geq 0\} \;\; and \;\; H^1(\ell) = \{W | g^1(\ell, W) > 0\}$$

where,

$$g^0(\ell, W) = -(X\beta + W), \; g^1(\ell, W) = X\beta + W \;\; and \;\; \ell = X\beta$$

The choice probabilities for the probit model are thus given by,

$$\begin{aligned} P(Y=0|X) &= h^0(\ell) = P(V \in H^0(\ell)) = P(-(X\beta + V) \geq 0) = \Phi(-\ell) = \Phi(-X\beta) \\ P(Y=1|X) &= h^1(\ell) = P(V \in H^1(\ell)) = P(X\beta + V > 0) = \Phi(\ell) = \Phi(X\beta) \end{aligned}$$

Since the standard normal distribution is strictly log-concave, Theorems 1 and 2 imply that both $P(Y=0|X)$ and $P(Y=1|X)$ are strictly log-concave in ℓ and hence strictly log-concave in $-X\beta$ and $X\beta$ respectively. Then, if there exists no linear subspace of $\mathbb{R}^k$ containing X with probability 1, both $P(Y=0|X)$ and $P(Y=1|X)$ are strictly log-concave in β as well.

2.1.2 Bivariate Ordered Response Probit Model

Consider the bivariate ordered response probit model with three outcomes (Greene and Hensher (2010)),

$$Y_1 = m \; if \; \alpha_m < X_1\beta_1 + V_1 \leq \alpha_{m+1} \quad and \quad Y_2 = n \; if \; \delta_n < X_2\beta_2 + V_2 \leq \delta_{n+1}$$

where $(m,n) \in \{0,1,2\}$, $\alpha_0 = \delta_0 = -\infty$, $\alpha_1 = \delta_1 = 0$ and $\alpha_3 = \delta_3 = \infty$, X_1 and X_2 are two $(1 \times k)$ vectors of regressors, β_1 and β_2 are two $(k \times 1)$ vectors of parameters of interest, α_2

[5]Pratt (1981) discusses strict log-concavity in the ordered response probit model.

and δ_2 are unknown threshold coefficients and $V = (V_1, V_2)' \in \mathbb{R}^2$ is the vector of unobserved heterogeneity, such that $V \perp (X_1, X_2)$ and $V \sim N(0, \Sigma)$, where Σ is known. Using the notation in Section 2 define as,

$$H^{mn}(\ell) = \left\{ W : \begin{array}{c} g_1^{mn}(\ell, W) > 0 \\ g_2^{mn}(\ell, W) \geq 0 \\ g_3^{mn}(\ell, W) > 0 \\ g_4^{mn}(\ell, W) \geq 0 \end{array} \right\}, \left\{ \begin{array}{c} g_1^{mn}(\ell, W) = X_1\beta_1 + W_1 - \alpha_m \\ g_2^{mn}(\ell, W) = \alpha_{m+1} - X_1\beta_1 - W_1 \\ g_3^{mn}(\ell, W) = X_2\beta_2 + W_2 - \delta_n \\ g_4^{mn}(\ell, W) = \delta_{n+1} - X_2\beta_2 - W_2 \end{array} \right\} \text{ and } \ell = \left\{ \begin{array}{c} \ell_1 = X_1\beta_1 - \alpha_m \\ \ell_2 = -(X_1\beta_1 - \alpha_{m+1}) \\ \ell_3 = X_2\beta_2 - \delta_n \\ \ell_4 = -(X_2\beta_2 - \delta_{n+1}) \end{array} \right\}$$

The choice probabilities for the bivariate ordered response probit model are thus given by,

$$P(Y_1 = m, Y_2 = n | X_1, X_2) = h^{mn}(\ell) = P(V \in H^{mn}(\ell))$$

Since $P(V \in H^{mn}(\ell))$ is the probability that the random vector V is contained in the convex polytope $H^{mn}(\ell)$, generated by a bivariate normal distribution with known Σ which is strictly log-concave, Theorems 1 and 2 imply that $P(Y_1 = m, Y_2 = n | X_1, X_2)$ for all (m, n) is strictly log-concave in ℓ. Then, if there exists no linear subspaces of $\mathbb{R}^k$ containing X_1 and X_2 with probability 1, $P(Y_1 = m, Y_2 = n | X_1, X_2)$ for all (m, n) is strictly log-concave in $\beta_1, \beta_2, \alpha_2, \delta_2$ as well.

3 Conclusion

Log-concave distributions cover a wide range of the most commonly used distributions in statistical and econometric models. This paper shows that in models where the multivariate random vector of unobservables follows a strict log-concave distribution, the probability that it is contained in a convex polytope is also strictly log-concave in the thresholds that define the half-space representation of the convex polytope. In linear index discrete response models with additively separable unobservables, where the response probabilities are given by the probability the multivariate vector of unobservables lies in a convex polytope, under the standard rank condition, this result implies that the likelihood function will be strictly log-concave in the parameters, as shown for example in Theorem 3 of Wedderburn (1976). Therefore, this result provides a sufficient condition for point-identification of the parameters, existence of a unique MLE and no other critical points, when for example, as discussed in Mäkeläinen et al. (1981), the gradient vector of a twice-continuously differentiable log-likelihood function vanishes in at least one point in the parameter space. This condition together with the strict log-concavity of the likelihood function[6], ensure that not only the log-likelihood has a unique global maxima and no other stationary points inside the parameter space, but also it has no local maxima on the boundary[7].

Furthermore, the strict log-concavity assumption is directly related to when algorithm based procedures, such as the EM algorithm[8], converge to a global optimum. For example, as shown in Wu (1983), Corollary 1, if the likelihood function is unimodal in the parameter space with a unique stationary point, then any EM sequence converges to the unique maximizer of the log-likelihood function, under a certain differentiability condition[9]. The strict log-concavity of the likelihood function in the parameters implies that it is also unimodal in the parameters [10]; thus, the EM algorithm is guaranteed to convergence to the unique MLE, when a unique stationary point exists.

References

Albert, A., Anderson, J., 1984. On the existence of maximum likelihood estimates in logistic regression models. Biometrika 71, 1–10.

Bagnoli, M., Bergstrom, T., 2005. Log-concave probability and its applications. Economic theory 26, 445–469.

[6]The negative definite Hessian matrix at every point in the parameter space in condition (ii) of Theorem 2.6 in Mäkeläinen et al. (1981) is a sufficient condition for the strict log-concavity of the likelihood function.

[7]Under for example complete or quasicomplete separation, discussed in Albert and Anderson (1984), the MLE is at infinity on the boundary of the parameter space. As also pointed out by Peters and Chesher (2000) in these two improper data configurations unique and finite maximum likelihood estimates do not exist.

[8]Lange (2013) highlights that "In the absence of concavity, there is also no guarantee that the EM algorithm will converge to the global maximum".

[9]McLachlan and Krishnan (2007) provide an extensive discussion of the theory, methodology and applications of the EM algorithm.

[10]See Dharmadhikari and Joag-Dev (1988), page 178.

Delle Site, P., Kilani, K., Gatta, V., Marcucci, E., de Palma, A., 2019. Estimation of consistent logit and probit models using best, worst and best–worst choices. Transportation Research Part B: Methodological 128, 87–106.

Demidenko, E., 2001. Computational aspects of probit model. Mathematical Communications 6, 233–247.

Dharmadhikari, S., Joag-Dev, K., 1988. Unimodality, convexity, and applications. New York: Academic Press.

Greene, W.H., Hensher, D.A., 2010. Modeling ordered choices: A primer. Cambridge University Press.

Lange, K., 2013. Optimization. volume 95. Springer Science & Business Media.

Lesaffre, E., Kaufmann, H., 1992. Existence and uniqueness of the maximum likelihood estimator for a multivariate probit model. Journal of the American Statistical Association 87, 805–811.

Mäkeläinen, T., Schmidt, K., Styan, G.P., 1981. On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. The Annals of Statistics , 758–767.

McCullagh, P., 1980. Regression models for ordinal data. Journal of the Royal Statistical Society. Series B (Methodological) 42, 109–142.

McLachlan, G.J., Krishnan, T., 2007. The EM algorithm and extensions. volume 382. John Wiley & Sons.

Orme, C.D., Ruud, P.A., 2002. On the uniqueness of the maximum likelihood estimator. Economics Letters 75, 209–217.

Peters, S., Chesher, A., 2000. On the use of enumeration for investigating the performance of hypothesis tests for economic models with a discrete response variable. Computational Economics 15, 273.

Pratt, J., 1981. Concavity of the log likelihood. Journal of the American Statistical Association 76, 103–106.

Prékopa, A., 1973. On logarithmic concave measures and functions. Acta Scientiarum Mathematicarum 34, 335–343.

Prékopa, A., 1980. Logarithmic concave measures and related topics, in: M.A.H. Dempster (Ed.), Stochastic Programming. Academic Press, pp. 63–82.

Prékopa, A., 2007. On the relationship between probabilistic constrained, disjunctive and multiobjective programming. RUTCOR Research Reports , 7–2007.

Prékopa, A., 2012. Multivariate value at risk and related topics. Annals of Operations Research 193, 49–69.

Silvapulle, M., 1981. On the existence of maximum likelihood estimators for the binomial response models. Journal of the Royal Statistical Society. Series B (Methodological) 43, 310–313.

Silvapulle, M., Burridge, J., 1986. Existence of maximum likelihood estimates in regression models for grouped and ungrouped data. Journal of the Royal Statistical Society. Series B (Methodological) 48, 100–106.

Wedderburn, R., 1976. On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. Biometrika 63, 27.

Wu, C.J., 1983. On the convergence properties of the em algorithm. The Annals of statistics , 95–103.